R E P O R T R E S U M E S

ED 002 971 24

THE EFFECT OF TELEVISION INSTRUCTION ON INDIVIDUAL LEARNING CURVES.

BY- STAKE, ROBERT E.
NEBRASKA UNIV., LINCOLN

REPORT NUMBER CRP-573 PUB DATE NOV 59
CONTRACT OEC-SAE-8190
EDRS PRICE MF-$0.09 HC-$1.68 42P.

DESCRIPTORS- *EDUCATIONAL TELEVISION, *TELEVISED INSTRUCTION, SPANISH, *ACHIEVEMENT GAINS, *COMPARATIVE ANALYSIS, LEARNING ACTIVITIES, *CONVENTIONAL INSTRUCTION, HIGH SCHOOL STUDENTS, VOCABULARY DEVELOPMENT, OPEN CIRCUIT TELEVISION, LINCOLN, NEBRASKA

THIS RESEARCH WAS DESIGNED TO TEST THE HYPOTHESIS THAT THERE IS NO DIFFERENCE BETWEEN FORMAL LEARNING CURVES OF STUDENTS TAUGHT SPANISH IN A TELEVISION-CORRESPONDENCE COURSE AND THOSE TAUGHT SPANISH IN THE CONVENTIONAL MANNER. AN EXPERIMENTAL GROUP OF OVER 20 STUDENTS IN 5 SMALL HIGH SCHOOLS WAS GIVEN INSTRUCTION IN SPANISH FOR 1 YEAR, ENTIRELY BY OPEN CIRCUIT TELEVISION. THREE CONTROL GROUPS (TAUGHT SPANISH IN THE CONVENTIONAL MANNER) WERE FORMED AS FOLLOWS--(1) OVER 30 STUDENTS FROM 3 SMALL HIGH SCHOOLS, (2) ABOUT 100 STUDENTS FROM A LARGE HIGH SCHOOL IN A PARTIALLY SPANISH-SPEAKING COMMUNITY, AND (3) APPROXIMATELY 15 STUDENTS IN A COLLEGE CLASS. ALL GROUPS RECEIVED THE SAME INSTRUCTIONAL PROGRAM, A 1-YEAR ELEMENTARY SPANISH COURSE. THE VOCABULARY ABILITY OF EACH GROUP WAS ASSESSED EVERY 2 WEEKS WITH A 15-MINUTE QUIZ. LEARNING CURVES WERE GENERATED ANALYTICALLY, USING A MODIFIED RATIONAL GYPERBOLA MODEL, BASED ON THE QUIZ RESULTS. SIGNIFICANT GROUP DIFFERENCES WERE NOTED WITH REGARD TO LEVEL OF ACHIEVEMENT. AT THE BEGINNING OF THE PROJECT, THE LARGE CONTROL GROUP WAS MORE PROFICIENT THAN EITHER THE SMALL SCHOOL CONTROL OR EXPERIMENTAL GROUPS (WHICH WERE ABOUT THE SAME). BY THE END OF THE FIRST SEMESTER OF WORK, THE SMALL HIGH SCHOOL CONTROL GROUP HAD PASSED THE LARGE SCHOOL CONTROL GROUP, BUT THE EXPERIMENTAL SUBJECTS WERE SIGNIFICANTLY BEHIND BOTH. THESE FIRST SEMESTER DIFFERENCES WERE MAGNIFIED EVEN MORE AT THE END OF THE SCHOOL YEAR. THE SIGNIFICANTLY POORER ACHIEVEMENT BY THE EXPERIMENTAL GROUP WAS BRIEFLY ASSESSED TO DETERMINE PROBABLE CAUSES. IT WAS APPARENT THAT MORE STUDY WAS NEEDED IN ORDER TO ESTABLISH A PROPER ROLE FOR EDUCATIONAL TELEVISION. (JH)

ED002971
REFERENCE COPY
COOPERATIVE RESEARCH PROGRAM
NO. 573

573

OFFICIAL RESEARCH FILE COPY

The Effect of Television Instruction on Individual Learning Curves

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Robert E. Stake

Department of Educational Psychology and Measurements

University of Nebraska

November 1959

Project No. 573(8190)

The research reported herein was performed pursuant to Contract #SAE-8190, United States Office of Education, Department of Health, Education and Welfare and the University of Nebraska. Reproduction of any part of this material, exclusive of copyrighted material, is permitted for any purpose of the United States Government.

Acknowledgements

The author is indebted to a large number of people and organizations and wishes to express his gratitude for their contributions:

to Gayle Childs, Niels Wodder, Leon Rottmann, and Charles Neidt for their advice and support,

to Lloyd Teale for his help in clarifying several of the objectives in teaching Spanish,

to the high school teachers, who will remain anonymous because of the comparative nature of the study, for their participation,

to Don Stake who equated the quizzes, to Jo Buck who scored the 120,000 items and typed the manuscript, and to Ben Shaw who helped with the computer programming, and to Nora Judd who prepared the tables,

to the Educational Testing Service for permission to use the Cooperative Spanish test items,

to the State Farm Mutual Insurance Company for permission to use their computational facilties and to the International Business Machines Company for numerous favors,

and to the U. S. Office of Education which gave financial support for the major phases of the experiment.

List of Tables

Table	Title	Page
1	Means of parameters and test scores for students instructed for the full year	14
2	Standard deviations of parameters and test scores for students instructed for the full year	17
3	Correlation matrix for parameters and test scores for students instructed for the full year	19
4	Signs of discrepancies between theoretical curves and actual plots of data for students instructed for the full year	26
5	Coefficients of correlation between curve parameters for first semester curves and for full year curves for students instructed for the full year	27

List of Figures

Figure	Title	Page
1	Group learning curves generated from mean parameters	21
2	Learning curves for four students	24

THE EFFECT OF TELEVISION INSTRUCTION ON INDIVIDUAL LEARNING CURVES

Table of Contents

	Page
The Problem	1
Previous Investigations	1
The Learning Task	2
The Students	4
The Testing	5
The Learning Curves	6
Available Curve Parameters	8
The Choice of Parameters	11
The Analytic Procedure	11
Experimental Design	12
The Results	13
Attrition	13
Computation	13
Group Characteristics	14
The Intercorrelations	18
Discussion	20
The Experiment	20
The TV Group	20
The Control Groups	22
The College Group	22
The Parameters	23
The Hyperbola	23

Table of Contents

(continued)

Page

The Significance of the Results of the Experiment 30

Recommendations 31

Summary 32

The Problem 32

The Procedure 32

The Findings 33

Conclusions 33

References 35

The Problem

There are many dimensions to achievement. Even in a simple task there are many different aspects to the performance. Furthermore, there are many different ways of attaining a high level of achievements. This study was undertaken to inspect some of the different paths that individual students take in learning, with particular attention to learning when all instruction is via television.

Previous Investigations. Many of the experiments comparing TV instruction to conventional instruction have shown that students taught by TV do as well as those taught in the conventional manner, whereas other have found some superiority in one procedure or the other. Carpenter & Greenhill, 1958; Herminghaus, 1958; Kumata, 1956; Macomber, 1957; Neidt & French, 1958; Rottmann, 1959; Tanner, 1957). The criterion usually has been performance on some type of final achievement test. In the better experiments an attempt has been made to control the pre-experiment training or level of achievement or intelligence of the students. Whether initial standing is controlled or not the results of these experiments have investigated only a single criterion or aspect of the learning situation. It is the purpose of this study to investigate several different aspects of the learning situation in which both TV and conventionally taught students are involved.

In other studies attention has been focussed on the difference between final test scores for students taught by TV and in the conventional manner. It is not known whether or not whatever improvement there is for the two groups is at the same rate and whether or not the TV instruction results in more "regular" learning. The situation in which TV is used today is a novel situation for the high school student. It might be that learning at the outset in this situation

tends to be relatively slow. On the other hand, with greater concentration of talent and with superior organizational and resource facilities, or just because of its novelty, the TV instruction might evoke more rapid learning at the outset of instruction. In either case, the initial superiority or inferiority might give way to a completely opposite trend at midyear or year's end. This study was undertaken because it was felt that the advantages and disadvantages of TV instruction could not properly be discussed until the week to week progress of individual students was analyzed.

Gain scores, the differences between initial and final achievement, have been shown to lack adequate reliability and validity as measurements of improvement. (Lord, 1956; Woodrow, 1946). Learning curves, which more fully portray the improvement in proficiency, can yield parameters associated with the tendency of the student to respond correctly at any stage of training, with the rapidity of his improvement, with the regularity of his improvement and with other aspects or dimensions of his learning. Mathematically, there is no end to the number of parameters that could be isolated and used to describe a student's learnings. It is possible, of course, to measure very carefully a lot of things which turn out to be no more than the artifacts of random or meaningless activity. Whether or not greatly detailed analysis and description is justified depends on the degree to which the descriptive parameters conform to or enhance our understandings of learning.

The Learning Task. There are many objectives to any course of study. To evaluate the progress of students completely, it is necessary to measure the success of the pursuit of each of the objectives during

the training. Certainly it is expected that the attainment of different objectives will procede in different fashions, but where there are interrelationships among the objectives, certain ones can be studied which will be, at least in part, representative of the learning as a whole. At any rate, study of some of the more accessible aspects of the training is a logical first step.

The learning task selected for investigation here was the recognition of elementary Spanish vocabulary, one of the aspects of learning in the usual high school Spanish class. Study of a language was chosen because it seems to be relatively homogeneous and cumulative learning activity, an activity with few necessary digressions, with a mimimum of the topicality found in such studies as civics, physics, and literature, and with a great dependence on and review of the material learned earlier in the course. In order that the learning be meaningfully described by a single learning curve, measurements must be made on the same or equivalent instruments throughout the training. A homogeneous activity was preferred because with such an activity more items of that one evaluative instrument are likely to contribute to valid discrimination at any point in the training. This is especially important when rather limited time is available for the testing.

It was recognized that in all foreign language classes the teacher intends to accomplish much more than to teach vocabulary. He usually intends to teach pronunciation, idiom, word derivation, grammar, and a bit about the literature of that tongue. Teachers emphasize these objectives differently. It is probably true that there was considerable difference among the teachers used in this

study as to the relative importance of the mastery of vocabulary, and they probably devoted different amounts of time and energy toward this goal. However, there was no reason to believe that vocabulary exposure with one medium of instruction would be systematically different than with the other medium. And because of the necessity of focussing attention on some aspect of the learning and because of the desirable features of vocabulary learning and assessment, no other objective of elementary Spanish training was investigated. Proficiency with vocabulary was the only element of achievement analyzed in this study.

The Students. The experimental group in this study was a group of students in several high schools who were enrolled in a TV-correspondence course in beginning high school Spanish. On 3 days a week for 2 semesters these students watched a TV lecture for 1 hour. A supervisor was present but this person was not qualified to provide instruction in Spanish and did not participate in the instruction. Assignments were mailed to a correspondence study center where they were corrected and returned. The high schools which these students attended were very small schools which did not provide Spanish in the usual curriculum.

There were two control groups, one a group of students in three small high schools and the other a group of students in four classes in one of the largest high schools in the state of Nebraska. This latter school is unique in that quite a few people in the community speak some Spanish. The Spanish instruction for these students was roughly equivalent to the type of instruction provided for the experimental group.

The experimental and control groups had been identified and the instruction for them was specified by the research plan of the Ford

Foundation Nebraska In-School Television-Correspondence Study Program. This program had provided the same TV-correspondence instruction in Spanish as reported herein the previous year. (Rottmann, 1959).

Another group was brought into this study not as a control group but to lend further understanding of the results. This was a group of college freshmen and sophomores taking a regular first course in Spanish at the college level. These college students were involved in more hours of classwork than the control groups, plus a laboratory experience in Spanish each week.

The Testing. Every other week a 15 minute vocabulary quiz was administered to the students in each of these groups. In order to make the measurements on successive weeks parallel without the scores being a mere indication of how well the quiz items had been learned, 9 parellel quizzes were devised. By special permission of Educational Testing Service, Princeton, New Jersey, items from Parts I and II, Forms N, O, and P, of the Cooperative Spanish Tests were used. Copies of the 9 different tests were distributed at random to the students at each administration. This assured that the testing for the entire group was equivalent at each administration and diminished the opportunities for cheating. The testing was done in the regular classroom and proctored by the regular supervisor or instructor.

Each quiz included 40 five choice multiple-choice items. The answer sheets were scored with a correction for omissions, the score being the number right plus one fifth of the number omitted. Reports of results (total scores) were distributed to the teachers but copies of the quizzes were not available for review. Because the quizzes were

found to be not equally difficult and therefore not quite parallel, after all testing for the year was completed tables were empirically devised so that the scores from different quizzes could be equated. The scores that were used for the learning curves were right scores, corrected for omissions and equated for test difficulty.

The Learning Curves. Mathematical analysis of learning curves depends upon the arbitrary selection of a model. There are a large number of models from which to choose: stochastic models, learning theory models, empirical models, rational models. The rational model is one derived from certain basic assumptions or definitions. For this study Thurstone's (1930) rational hyperbola was chosen as the model. The equation for this hyperbola is as follows:

$$\left(\frac{g}{c} - w\right)\left(\frac{h}{c} + r\right) = \frac{g}{c}\frac{h}{c} \qquad (1)$$

where

r is the cumulative number of right reponses,

w is the cumulative number of wrong responses, (r + w = 40),

g is the initial strength of the tendency to respond correctly,

h is the initial strength of the tendency to respond incorrectly,

c is the change in habit strength following any response.

The basic assumptions for this model are that right and wrong responses are made in the ratio of the strengths of the two tendencies, that the strength of the tendency to respond correctly increases by a constant amount as a result of and only as a result of knowledge that a correct response was made and that the strength of the tendency to respond incorrectly decreases by the same constant amount as a result of and only as a result of knowledge that an incorrect response

was made. These assumptions yield the two parameter, hyperbolic relationship given in (1).

Whether or not the assumptions just mentioned are particularly appropriate for the learning of Spanish vocabulary, a hyperbolic relationship has several desirable features. It is assumed that all pupils could eventually learn the vocabulary perfectly, therefore, all learning curves could be expected to head toward errorless performance i.e., a horizontal asymptote when cumulative errors are plotted against cumulative successes. The Thurstone hyperbola approaches a horizontal asymptote. It seems that the model to be used should be one which can be used in a situation where all the learning occurs at once. Although sudden insight on 40 questions is unlikely, the hyperbola model is one of the few models that can completely describe such a situation. Perhaps a more important consideration is that the hyperbolic model is quite appropriate for describing a situation in which the performance continues with successes occurring at a chance level for a while before learning is manifested. The Thurstone hyperbola does handle this situation if a slight modification is made to account for different probabilities of chance success. The modification is as follows. The observed number of right responses (R) is the true right responses (r) augemented by the guessed right responses,

$$R = r + pw \qquad (2)$$

where p is the probability of making a right response by chance.

The number of observed wrong responses (W) is given by

$$W = w - pw = qw \qquad (3)$$

where $q = 1 - p$, the probability of making a wrong response by

chance. Solving for r and w and substituting (2) and (3) into (1) we have

$$\left(\frac{g}{c} - \frac{W}{q}\right)\left(\frac{h}{c} + R - \frac{p}{q}W\right) = \frac{g}{c}\frac{h}{c} \qquad (4)$$

which is the formula for a hyperbola having one asymptote extending horizontally and the other asymptote, a line representing chance performance, having the slope q/p. (W is plotted on R). In formula (4) there are two independent parameters, g/c and h/c, the arbitrary constants, p and q , and the variables, R and W .

Available Curve Parameters. The use of hyperbola (4) is based on just two independent curve parameters. Different combinations of these parameters can be used for certain descriptive purposes, but once two parameters have been specified, other parameters provide basically no additional information. The parameters, g/c and h/c , can be used as defined: g/c is the initial strength, in increment units, of the student's tendency to respond correctly and h/c is the initial strength, in increment units, of the student's tendency to respond incorrectly.

Other parameters seem to be more meaningful to most persons in education. The initial tendency to respond correctly i.e., the probability a right response will be made, appears to be a desirable parameter. This parameter is the slope of the curve, if W is plotted on R , which can be determined for each individual on any task at any time during training by

$$\frac{1}{\frac{p}{q} + \frac{\frac{h}{c}\frac{g}{c}q}{\left(\frac{g}{c}q - W\right)^2}} \qquad (5)$$

where W is the total number of wrong responses that have been recorded at that time.

Since p = 1/5 and q = 4/5 in a 5 choice multiple choice situation the general slope parameter for this study can be written

$$\frac{1}{\frac{1}{4} + \frac{\frac{4}{5}\frac{h}{c}\frac{g}{c}}{\left(\frac{4}{5}\frac{g}{c} - W\right)^2}} \quad . \qquad (6)$$

Since W = O at the outset of training the initial slope of the curve, i.e., the initial tendency to respond correctly, is

$$\frac{1}{\frac{1}{4} + \frac{5}{4}\frac{h}{g}} \quad . \qquad (7)$$

It is clear in (6) that the expression approaches zero as W approaches (4/5)(g/c). The slope is zero when perfect performance is reached. Therefore, the W intercept of the horizontal asymptote for each individual learning curve is

$$\frac{4}{5}\frac{g}{c} \quad . \qquad (8)$$

The tendency to respond correctly at the conclusion of training can be obtained from (6) after g/c and h/c are determined by substituting the grand total of errors for W.

The rate of learning can be defined as the radius of curvature of the learning curve. This curvature varies along the curve but can be used as a parameter if a certain reference point is specified. Comparison at any other reference point would yield identical ordering of the students as to rate of learning. The general expression for the radius of curvature here is

$$\frac{\frac{8}{5}\frac{h}{c}\frac{g}{c}}{\left(\frac{4}{5}\frac{g}{c} - W\right)^3} \qquad (9)$$

The curvature will reach a maximum when the hyperbola intersects its major axis. This occurs when the slope is parallel to a line that bisects the acute angle of the asymptotes; when, in this analysis, the curve slope equals .781 . The number of wrong responses that a student made before his curve attained such a slope can be determined and substituted into (9).

In every use of hyperbola (4) the maximum radius of curvature is found to be

$$K\sqrt{\frac{h}{c}\frac{g}{c}} \qquad (10)$$

where K is a constant. In this investigation K = .427. With a different value of K, (10) is also the equation for the distance from the point of maximum curvature to the intersection of the asymptotes, i.e., the semitransverse axis. In this investigation, the K for this parameter is .698. Ignoring the constant, the maximum curvature can be represented by

$$\sqrt{\frac{h}{c}\frac{g}{c}} \quad . \qquad (11)$$

This curvature parameter is identical to one recommended by Wiley and Wiley (1937) as the best measure of learning ability.

The regularity of a student's performance can be indicated by the amount the observed performance departs from the smooth theoretical performance--provided, of course, that the theoretical performance is appropriately determined. These departures or deviations are conventionally represented by a standard error, the square root of the mean

of the squared deviations. This standard error is a goodness of fit parameter rather than a curve parameter.

The Choice of Parameters. Although only two curve parameters can be considered completely independent when the modified Thurstone hyperbola is used, it was decided that three curve parameters would be reported: the initial slope parameter of equation (7), the final slope parameter of equation (6), and the curvature parameter of equation (11). These parameters were intended to be indices, respectively, of the student's initial tendency to respond correctly, final tendency to respond correctly, and rate of learning. (It should be noted that these three parameters are on inverted scales: the higher the score, the poorer the performance.) In addition, the goodness of fit parameter was to be reported as an index of the regularity of the student's performance.

The Analytic Procedure. The solution for all parameters can be facilitated by a solution for the two parameters, (h/c) and (g/c) , of equation (4). There are no direct solutions for any of these parameters, but indirect, iterative solutions such as the Method of False Position can be used. These are tedious, however, and it was found in Stake's analysis of laboratory learning curves using the modified Thurstone hyperbola (1958) that passing the hyperbola through the origin and two selected points (thus permitting a direct solution for the parameters) yielded acceptable results. The standard errors he obtained were reasonably small for a very smooth performance and the theoretical curve, when plotted, appeared to represent the actual data reasonably well. The same procedure was followed in this in-

vestigation. The hyperbola was forced to pass through a centroid of five (or sometimes three) performances midway through the training and through a second centroid of the final five (or three) performances. Since each centroid has R and W coordinates which must satisfy equation (4), these values and the known values of p and q can be substituted into equation (4) to yield two equations with two unknowns. The parameters are determined by solving the simultaneous equations.

Experimental Design. The design of the experiment called for instruction in Spanish, via TV for the experimental group and via conventional procedures for the control groups, throughout the school year 1958-1959. The progress was to be assessed with a 15 minute quiz every two weeks and with a one hour standardized examination at the end of each semester.

The four parameters mentioned in the previous paragraph were to be used to test the hypothesis that there is no difference between TV-taught and conventionally taught groups with regard to early achievement, final achievement, rate of learning and regularity of performance. In addition, computation of the coefficients of correlation between the parameters was scheduled. Further, the parameters were to be compared to the achievement scores earned at midyear and year's end, i.e., the conventional measurement of achievement, and to intelligence test (SCAT) scores.

It was considered desirable to obtain other test scores from the students, such as curve parameters for non-academic learning and scores of ability to listen effectively. Several of the participating schools were adamant, however, in their refusal to schedule further testing.

The Results

Instruction and testing were carried out as planned. At the end of the year it was discovered that the Cooperative Spanish tests had been used as semester examinations as well as the source of items for the quizzes. A different standardized examination would have been preferable.

Attrition. The loss of students, particularly from the control group, was discourging. When the research was planned it was expected that there would be over one hundred in the experimental group. Enrollments were considerably down from the previous year and there were many drop-outs. The attrition is apparent in the following table:

Group	Students Enrolled	Students Finished 1/2 year	Students Finished 1 year
Experimental	45	37	23
Small Control	33	32	31
Large Control	105	104	96
College	50	37	14

Computation. The parameters and intercorrelations were determined by the use of a special IBM 650 high speed computer program. A straight line was fitted to those cases where no learning occurred. The "curvature" of a straight line was arbitrarily set at 50, a figure just larger than the curvature score of the obtained individual curve showing the least learning. This curvature parameter could vary, then, from 0 to 50. The slope parameters could vary from 0 to 4. The fit parameter could vary from 0 to 60. The achievement tests, like the

quizzes, were scored to range from 0 to 40 with a chance score being 8.

Group Characteristics. The mean parameters for the four groups of students are shown in Table 1. The means for the control groups and college group have been compared with the experimental group means with conclusions based on the distribution of t with 22 degrees of freedom for the control groups and 13 degrees for the college group.

It is seen that the mean initial slope of curves (the tendency to respond incorrectly early in training) for the experimental group is significantly steeper than for the large school control and college groups. The experimental and small school control groups do not differ significantly in this regard. Considering all four groups it appears that initial vocabulary test performance of the TV group was inferior to that of the other groups.

The final slope (the tendency to respond incorrectly in the final phase of training) of the experimental group curves tends to be steeper than the final slope of the curves for the other students. The difference between experimental and small school control is significant at this stage although it was not at the outset. The difference between experimental and large school control groups is not significant at the termination of training although it had been at the outset. The college group is very far ahead of all the high school groups at the end of the year of training.

The experimental group did not differ significantly from the other three groups with regard to hyperbola curvature. It was noted, with some surprise, that the college group curves tended to have the largest radii, i.e., the smallest overall rate of improvement. Had this difference been significant it would have led to the conclusion that when the high school

Table 1

Means of parameters and test scores for students instructed for the full year

	Group			
	Experimental	Small Control	Large Control	College
Parameter				
Initial Slope	3.12	2.96	2.74*	1.46***
Final Slope	1.12	0.64*	0.96	0.24***
Curvature	16.36	14.15	17.53	18.67
Fit	5.06	5.77	4.63	5.25
Test Score				
Sem Acht	13.09	16.10*	15.69*	
Year Acht	14.04	18.94**	18.69**	
IQ	69.65	65.55	67.14	
N	23	31	96	14

Note--Asterisks indicate the significance of the difference between the marked mean and the corresponding mean for the experimental group. For * $\alpha = .05$, for ** $\alpha = .01$. and for *** $\alpha = .001$.

students attained the mastery of vocabulary already attained by the college students the high school students would then be improving their vocabularies at a faster rate than the college students did.

The four groups were found to be quite homogeneous with regard to the goodness of fit characteristic, i.e., regularity of performance.

Both control groups were significantly better than the experimental group on the hour achievement tests administered at the end of the first semester. Performance on a parallel test at the end of the second semester showed that then the difference between the experimental group and the control groups was even more significant, statistically speaking.

The differences among performances on the intelligence test were not statistically significant. The direction of the differences found was such that they could in no reasonable way account for the differences among learning curves. It was decided that it was not necessary to control for intelligence when comparing the groups, even though there is a considerable correlation between intelligence scores and some of the parameters.

The standard deviations of the groups with regard to curve parameters and test scores are shown in Table 2. With regard to initial slope the large control group was heterogeneous, the other groups were rather homogeneous. The small control group remained fairly homogeneous with regard to tendency to respond correctly during the training whereas these scores for the college student curves became much more homogeneous by the time training was terminited. The initial slopes for the experimental group were more homogeneous than the final slopes. (Interpretations of these variabilities should be made with caution.

Table 2

Standard deviations of parameters and test scores for students instructed for the full year

	Group			
	Experimental	Small Control	Large Control	College
Parameter				
Initial Slope	.50	.54	.88**	.44
Final Slope	.87	.49**	.70	.11**
Curvature	8.91	3.30**	9.66	4.73*
Fit	2.87	2.96	2.54	1.68*
Test Score				
Sem Acht	3.88	3.91	6.28**	
Year Acht	4.23	5.47	7.20**	
IQ	16.68	15.67	18.96	
N	23	31	96	14

Note--Asterisks indicate the significance of the ratio between the marked standard deviation and the standard deviation of the experimental group. For * $\alpha = .05$ and for ** $\alpha = .01$.

Slope scores are not equal interval scale scores. The standard deviation of slopes for a heterogeneous group of highly proficient students would probably be smaller than the standard deviation of slopes for a homogeneous group of poor students.)

The curvature parameters were about equally scattered for the experimental and large control group but the small control group was significantly more homogeneous. The homogeneity of fit parameters was essentially the same for the three high school groups.

The small control group and experimental groups were about equally homogeneous on both achievement tests. The large control group was relatively heterogeneous on both. All three groups were about equally homogeneous with regard to intelligence scores.

The Intercorrelations. The Pearson product moment coefficients are shown in Table 3. Coefficients as large or larger than .41 for the experimental group, .35 for the small school control group, .20 for the large school control group and .53 for the college group are significantly different from zero correlation at the 95% level of confidence. Discussion of the information in Table 3 is presented in the next chapter.

Table 3

Correlation matrix for parameters and test scores for students instructed for the full year

	Group	A. Initial Slope	B. Final Slope	C. Curva-ture	D. Std Error	E. Midyr Acht Test	F. Final Acht Test	G. IQ
	Exp		.11	-.63	.17	.36	.58	.80
A.	Sm C		.15	-.77	.52	.69	.71	.60
	L C		.30	-.61	.18	.86	.74	-.01
	Col		-.16	-.61	.49			
	Exp			.52	-.64	.39	.66	.16
B.	Sm C			.47	-.43	.60	.40	.32
	L C			.31	-.49	.48	.50	.22
	Col			.87	.00			
	Exp				-.53	.20	.05	-.50
C.	Sm C				-.67	-.18	-.29	-.26
	L C				-.28	-.38	-.25	.20
	Col				-.22			
	Exp					-.16	-.08	.29
D.	Sm C					.11	.45	.18
	L C					.07	.08	.08
	Exp						.56	.22
E.	Sm C						.81	.56
	L C						.78	.06
	Exp							.56
F.	Sm C							.53
	L C							.08

Note--Parameters inverted to make high score the 'good' performance.

Discussion

The Experiment. Conclusions from this study about the effectiveness of TV instruction for teaching elementary Spanish vocabulary must be set forth with reservation because the experimental group was small and because there was no attempt made to equate the instructors or educational objectives. The instructors may have differed considerably with regard to their abilities to teach. Rottmann (1959) found that this TV teacher on the same instructional program the previous year was rated as less effective by the TV students than the classroom teachers of Spanish who were rated by control group students. Perhaps more important than that is the fact that the teachers differed in the amount of attention they gave to the general vocabulary which the quizzes sampled. Furthermore, some of the teachers or supervisors were so upset by the difficulty of the quizzes that although they were urged not to, they discussed some quiz items with the class. Still, it is felt that the obtained curves of learning are representative of the achievement of the students in learning general Spanish vocabulary. The foregoing ramarks were made to indicate that it is more difficult here than it would be in a well controlled experiment to account for the differences that are found. The following discussion contrasts the TV and conventionally taught group but implicitly recognizes that the contrast might be due to extraneous factors.

The TV Group. The TV group performances, as a whole, were poorer than those of the control groups. They trailed at outset, they trailed at midyear, and they trailed at the finish. ~~The curves of Figure 1 reveal this graphically.~~ The differences between intelligence and learning rates between TV and conventionally taught groups were not significant. The

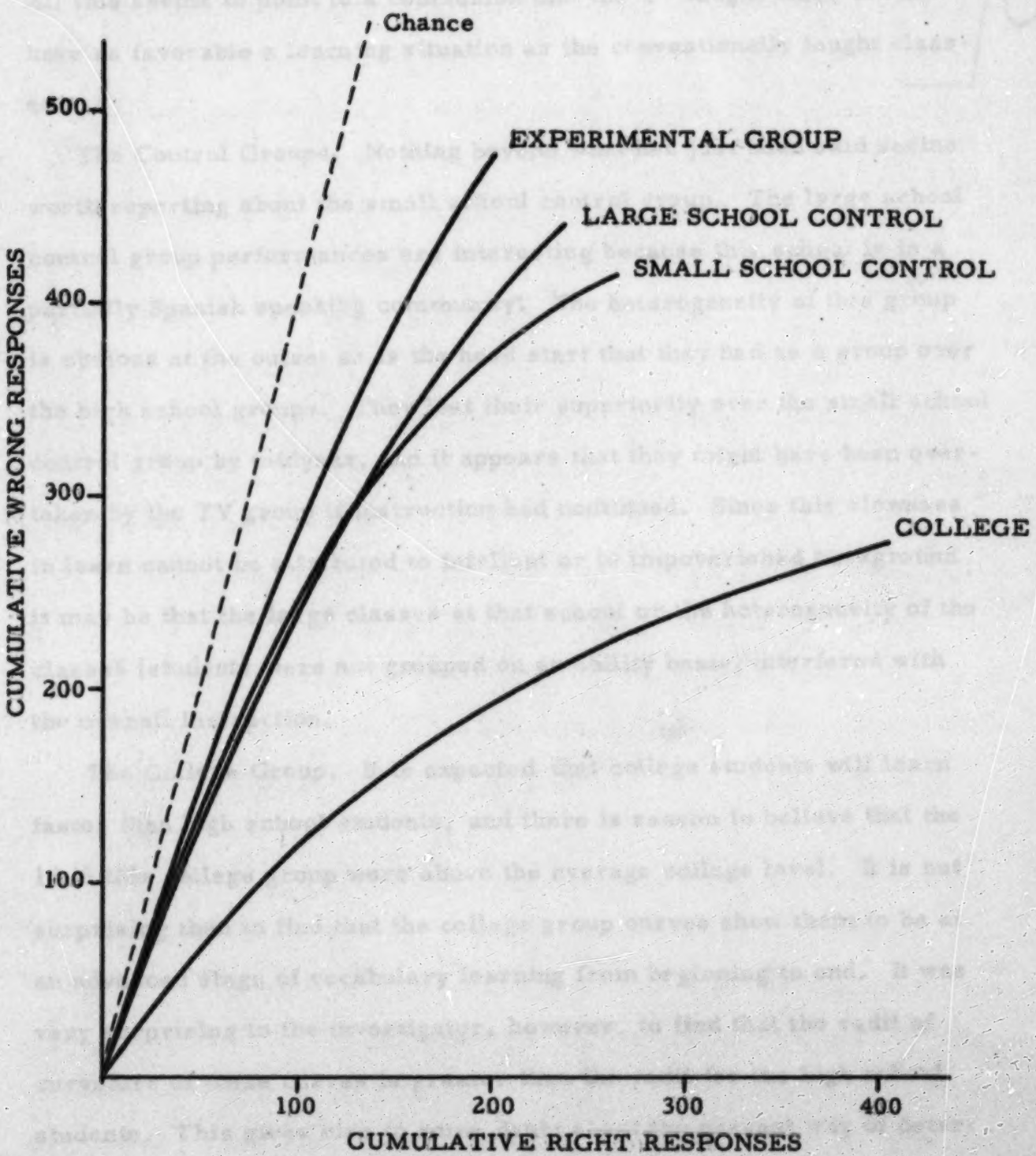


Fig. 1. Group learning curves generated from mean parameters.

differences between regularity of performance were also not significant. All this seems to point to a conclusion that the TV taught class did not have as favorable a learning situation as the conventionally taught classes.

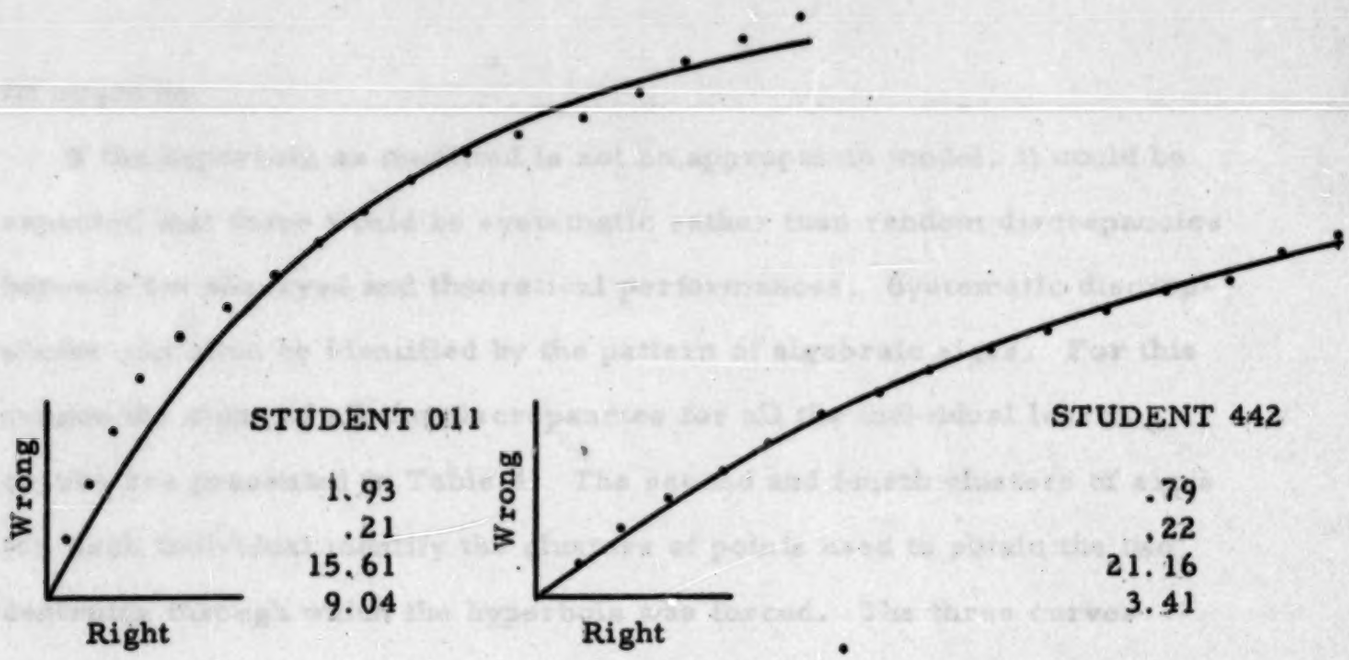
The Control Groups. Nothing beyond what has just been said seems worth reporting about the small school control group. The large school control group performances are interesting because this school is in a partially Spanish speaking community. The heterogeneity of this group is obvious at the outset as is the head start that they had as a group over the high school groups. They lost their superiority over the small school control group by midyear, and it appears that they might have been overtaken by the TV group if instruction had continued. Since this slowness to learn cannot be attributed to intellect or to impoverished background it may be that the large classes at that school or the heterogeneity of the classes (students were not grouped on an ability basis) interfered with the overall instruction.

The College Group. It is expected that college students will learn faster than high school students, and there is reason to believe that the 14 in this college group were above the average college level. It is not surprising then to find that the college group curves show them to be at an advanced stage of vocabulary learning from beginning to end. It was very surprising to the investigator, however, to find that the radii of curvature of these curves is greater than the radii for the high school students. This gives rise to some doubt about the present way of determining the curvature parameters. This will be discussed in a later section.

The Parameters. A number of conclusions about the parameters are based on Table 3, the table of intercorrelations. One such conclusion is that the two slope parameters are well chosen as descriptive indicants of learning here because there is very little correlation between them. This low correlation means that the persons who succeded most in learning vocabulary during the year were not noticeably better at the beginning. How well a student performed early in training seems largely based upon his intelligence. His tendency to respond correctly late in the year seems to be pretty much unrelated to his intelligence. Test scores obtained late in the year are fairly well related to IQ but final slope parameters are not. The midyear and final test scores appear to be more a function of intelligence than the final slope parameter is.

The fit or regularity of learning parameter is apparently related only to the final slope parameter and to the curvature parameter, and then only for the high school students. According to this relationship the students who had high learning rates and a strong tendency to respond correctly at the end of the year tended somewhat to be more irregular in their performances.

The Hyperbola. The hyperbola has desirable theoretical features and it seems to be powerful enough to fit a wide variety of performances well enough to satisfy visual scrutiny. In Figure 2 several individual performances are plotted and the hyperbola fitted. The standard errors for fitting smooth performances are quite low. Better fits, are possible, of course, with regard to the least squares criterion, but there is question as to whether the extra work involved in improvement in fit is worth the effort required to get it. So far it has not been determined how much better the

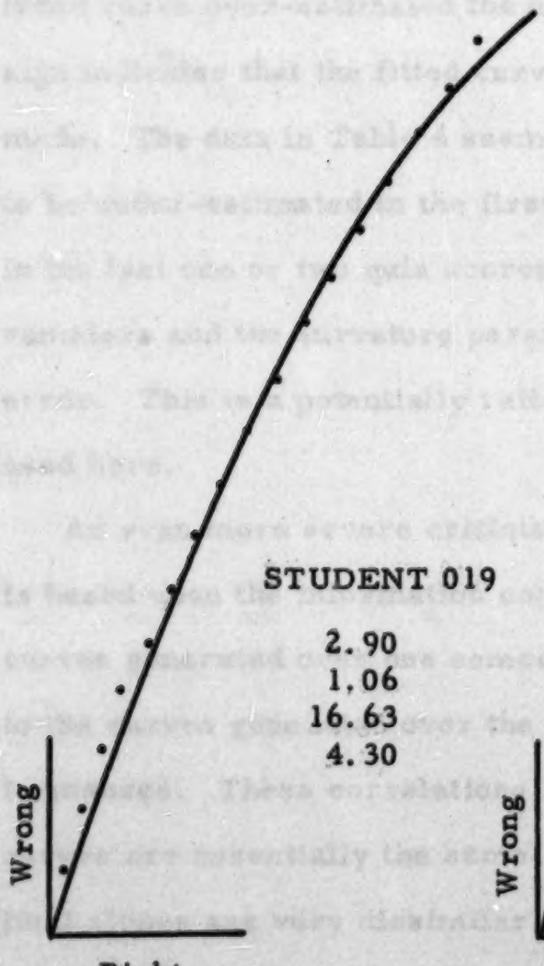


STUDENT 011

	1.93
	.21
	15.61
	9.04

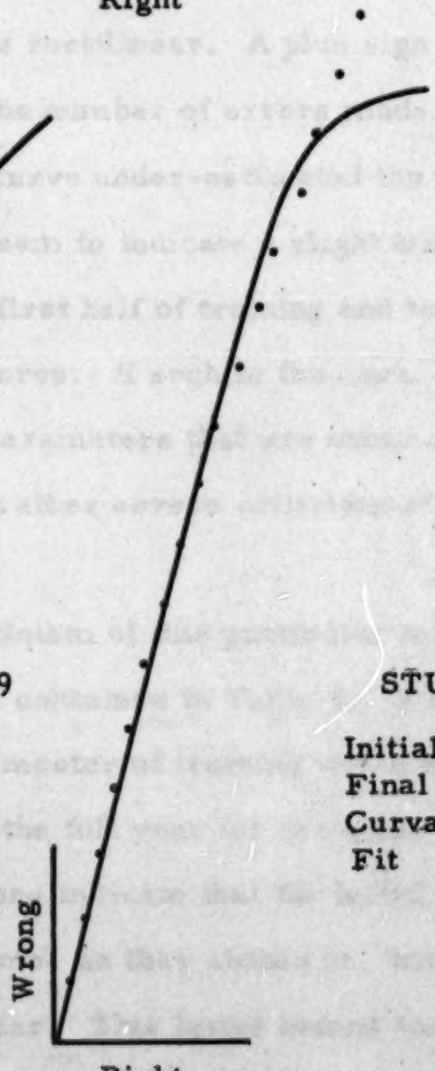
STUDENT 442

	.79
	.22
	21.16
	3.41



STUDENT 019

	2.90
	1.06
	16.63
	4.30



STUDENT 001

Initial Slope	3.96
Final Slope	.09
Curvature	5.46
Fit	14.85

Fig. 2: Learning curves for four students

fit might be.

If the hyperbola as modified is not an appropriate model, it could be expected that there would be systematic rather than random discrepancies between the observed and theoretical performances. Systematic discrepancies can often be identified by the pattern of algebraic signs. For this reason the signs of all the discrepancies for all the individual learning curves are presented in Table 4. The second and fourth clusters of signs for each individual identify the clusters of points used to obtain the two centroids through which the hyperbola was forced. The three curves identified with an asterisk are rectilinear. A plus sign indicates that the fitted curve over-estimated the number of errors made, whereas a minus sign indicates that the fitted curve under-estimated the number of errors made. The data in Table 4 seem to indicate a slight tendency for errors to be under-estimated in the first half of training and to be under-estimated in the last one or two quiz scores. If such is the case, the final slope parameters and the curvature parameters that are obtained are somewhat in error. This is a potentially rather severe criticism of the hyperbola as used here.

An even more severe criticism of this particular mathematical model is based upon the information contained in Table 5. It is expected that the curves generated over one semester of training would be rather similar to the curves generated over the full year for the same but extended performances. These correlations indicate that the initial slopes of the two curves are essentially the same, as they should be, but the curvature and final slopes are very dissimilar. This latter means that the curvature and final slope parameters are not very reliable general indices or that a

Table 4

Signs of Discrepancies between Theoretical Curves and Actual Plots of Data for Students Instructed for the Full Year

Experimental Group

001 +++-- --+++ ++ ++---
002 +---- -++++ ++ +--++
003 ----- ---++ ++ +----
004 +---- +++-- ++ ++---
005 ----- ----+ ++ +++--
006 ----- ---++ ++ ++--+
007 ----- ----+ -+ ++---
008 ----- ---++ ++ ++---
009 -++-+ --+-+ +- ++---
010 ----- +--+- -- ----+
011 ----- --+-+ -+ ++---
012 +++-+ +++-- -- -+---
013 ----+ +--+- -- +++--
014 -++-- ----+ -+ ++---
015 ++-++ ++-+- +- ++---
016 +++-- --+-+ ++ --++-
017 +-+-+ -+++- -- -++--
018 ----- +--++ +- ----+
019 ----- --+-+ ++ ++---
020 ----- ---++ ++ ++---
021 +++-- --+++ -- ++---
022 ----- -+--+ ++ +++--
*023 -+++-+++++--+-+-+

Small School Control

201 --++- +---+ -+ ++---
202 ----- --+++ ++ +--+-
203 ----- ----+ ++ ++---
204 +---- ----+ ++ ++---
205 --++- +----+ ++ ---++
206 -++++ ----+ ++ ++---
207 -+--- ---++ ++ +++--
208 ----- --+++ ++ ++---
209 +++++ -+--+ ++ ++---
210 ----- ----+ ++ ++---
211 ++++- --+-- ++ +++--
212 +++++ --+++ ++ +++--
213 ++++ ++--- -- -----
214 -+--+ --++- -- -++++
215 ----- --+-- +- --+--
216 +++-- ---++ ++ ++---
217 ----- --+++ ++ -----
218 +++++ ++--- -- ---++
219 ----- ---++ ++ ++---
220 +++++ +-++- -- -+---
221 ----- --+-- -- ++---
222 +---- ----+ ++ ++---
223 -+--- ----- ++ ++---
224 ----- +++-- -- -++--
225 ---++ ---++ ++ -----
226 --++- ---++ ++ --++-
227 --+++ +++-- +- -++--
228 ---- -++-- + -++--
229 ----- ---++ -- -+---
230 ----- ----+ +- ---++
231 -++-- ---++ +- -----

Large School Control

401 ++++ ----+ + ++---
402 ---- --+++ + +-++-
403 ---- ---++ - +----
404 -++- --++- + -----
405 -+++ ++--- + +----
406 +---- -++ +++ ++-
407 +++++ --+ +++ +--
408 ---- -+--- + +++--
409 ---+ +--+- - +---+
410 ---- --++- - ----+
411 +++-- ++- --+ ++-
412 ---- ---++ + ++---
413 +--- ---++ + +++--
414 -++- --++- - +++--
415 +++- ++--- - ++---
416 +--- ++--- - --++-
*417 -+++++-++-+----
418 ---+ +-++- + +--+-
419 ---- -+--- + ---+-
420 -+++ ++--- - -+---
421 ---+ +--+- - --+++
422 ++-- ---++ + +++--
423 ++++ +++-- - +++--
424 ---- ---++ + ++---
425 +++- ----+ - -++--
426 +++- ----+ + -----
428 ---- -+++- + ++---
429 +++- ---++ + +++--
430 ---- +---+ + +----
431 ----- --+ +++ ++-
432 ++-- -+++- + +-+--
433 ++++ ++--- + +----
434 -+++ ++--- - +----
435 --+- --++- + +----
436 ---++ -++ +++ --+
437 +++++ +-- +-+ +--
438 --++ ---+- + ++---
439 ++--- ++- --+ +--
440 +--- --+++ - ----+
441 +++- ---++ + ++---
442 ---- ---++ + +++--
443 ---- ---++ + +++--
444 ++++ --++- - ++---
445 ---- ----- + +++--
446 ++++ ++--- - --+-+
447 -+++ ----+ + ++---
448 ---- --++- - -++--
449 ++-- --++- + ++---
450 ++++ +++-- - --+++
451 ---- ----+ + +++--
452 --+- --+++ - ++---
453 ---- ---++ + ++---
454 ---+ -+++- - ---++
455 +--- ---++ - -+---
456 -+++ ----- - -----

457 +-++ ----- + +----
458 ---+ +---- + +++--
459 ---- ---++ + +++--
460 ++++ ---++ - ++---
461 ++-+ +---+ + ++---
462 ---- ---++ + ++---
463 ---- --+++ - ++---
464 ---- ----+ + ++---
465 ---+ ----+ + --+--
466 ---- ---++ + +++--
467 --+- ++--- - --+--
468 ++++ +-+-- - ----+
469 ++-- ---++ + +++--
470 ---+ --++- - --++-
471 -+++ +-++- - -+---
472 ++++ ---++ + ++--+
473 ---- --++- - ++---
474 +++- ---++ - --+-+
475 ---- ---++ + ---++
476 ---- ---++ + +++--
477 ++-+ ----+ + ++---
478 ---- --++- - ---++
479 -+++ ++--- - --+++
480 ---- --+++ + ++-+-
481 ---+- -++ +-- +--
482 ---- ---++ + +++--
483 ---- ----- + --+--
484 +-+- ---++ + +++--
485 ---- --++- - ---+-
486 -++- ---++ + +++--
487 -+++ +++-- - -+---
488 --++ +---- + ++---
489 ---- ---+- - -----
490 -+++ +++-- + +-+-+
491 ---- ---++ + ++---
492 ++++ +---+ + ++---
*493 ------+++-----++
494 ---- --+++ - ++--+
495 --++ +---- - ---++
496 ---- ---++ + ++---
497 ---- --+++ - -++-+

College Group

601 ----- ---++ ++ -----
602 ----- ---++ -+ +---+
603 ----- +---- -- -----
604 ----- --++- ++ +++--
605 ----- ----- -- +++--
606 ---- -++ +++ --+
607 ++--- --++- -- --+++
608 ----- --+-- -- -+++-
609 ----- ---++ -+ ++---
610 ----- --+++ ++ ++---
611 --+-- --++- -- -+---
612 ----- ++--- -- ---++
614 ----- ---++ -- --+++
615 +++++ -+--- -- +++--

Table 5

Coefficients of correlation between curve parameters for first semester curves and for full year curves for students instructed for the full year

Group	N	Initial Slope	Final Slope	Curvature
Experimental	23	.83	.31	-.16
Small Control	31	.84	.14	.44
Large Control	96	.95	.37	.24
College	14	.93	.35	.70

mere pair of independent curve parameters are quite inadequate for describing such performances.

All in all, the treatment of these performances by the modified Thurstone hyperbola seems to be less than completely satisfactory. The curvature parameters seem to be too highly related to initial performance, too independent of intelligence, and too unreliable. The final slope parameters seem to be somewhat over-estimated.

Study of certain cases such as that of student 001 in Figure 2 reveals a possible cause of some of these shortcomings. A few performances are about at the chance level for over twenty weeks. This means that the first centroid is about on the line of chance performance. If the centroid falls on that line and if any improvement at all is made in the second semester under the techniques used in this study a completely insightful learning curve is generated. Whenever the first centroid is close to the line of chance the student is likely to be given credit for more learning than the plot seems to justify, i.e., the curvature and final slope parameters will be too low. In allowing for complete insight even in unlikely situations and in basing the curve on just two centroid points, the validity of these two parameters may be jeopardized.

There seem to be two possible ways to improve matters. One is to attempt to modify the present analysis further and the other is to try out other models and compare results. A possible modification would be to solve for the parameters by using two slopes rather than the two centroids. It might help to do away with the restriction that the hyperbola pass through the origin. This would introduce a third parameter, thereby making the solution more difficult but permitting more accurate measurement of the curvature and final slope. Another modification to make it a three parameter model would be to generalize the hyperbola as Gulliksen (1934) has

done. Modification of the existing technique to allow the parameters to converge on the values associated with a least squares fit is still another possible remedy. With regard to other models, it seems that fitting a general polynomial to the same data would give a better indication of the appropriateness of the hyperbola. A direct least squares solution for parameters is possible when the general polynomial is used.

There is reason to question the validity of some of the parameters obtained here but there is no doubt in the mind of the investigator that these analytic procedures are potentially very useful for the evaluation of learning. Even when the criterion is fairly specific, such as in this study, there are many different facets to the performance which are worthy of attention. The results of the analysis here seem to point toward more complex, not simpler, methodology.

The Significance of the Results of the Experiment. It would seem that the conclusions stated earlier in this chapter are considerably vitiated by the conclusions about the parameters as just stated. It is true that a comparision of scores is not meaningful if the scores are not valid, but the principle parameters in this experiment are not invalid as group indices. There are somewhat distorted, but, in the main, they probably reflect the situation reasonably well. The results relating to the slope parameters merit serious consideration. The results relating to the curvature parameters should be validated before being accepted. There were no significant differences found between groups regarding curvatures and standard errors so the conclusions drawn earlier are not likely to be misleading.

In spite of the questions regarding the parameters the investigator

feels that a significant and meaningful difference has been shown between the TV taught and conventionally taught high school students. Because the TV students performed as well as the others on an intelligence test the investigator feels that the instructional situation for the TV group was relatively less conducive to learning elementary Spanish vocabulary.

Recommendations. It is believed that the data collected should be analyzed further to substantiate the findings presented in this report and to clarify the meaning of the curvature parameter. It is recommended that other models be used on these data to compare the merits of different models. It is recommended that there be more comparisions of learning curves for TV and conventionally taught students in situations where the instruction is equated in other respects. Furthermore, it is recommended that learning curve parameters, when refined, be subjected to factor analysis as the basis for clearer definition and more effective implementation.

Summary

The Problem. This research was designed to test the hypothesis that there is no difference between formal learning curves for students taught Spanish in a television-correspondence course and those taught Spanish in the conventional manner. Attention was to be drawn to the initial slope of each curve as indication of the student's tendency to respond correctly early in training, to the final slope of each curve as indication of the student's tendency to respond correctly at the termination of training, to the maximum curvature as indication of the student's learning rate, and to the size of the discrepancies between theoretical curve and actual plot of data as indication of the student's regularity of performance. The intercorrelations among these parameters and achievement and intelligence tests were also to be determined.

The Procedure. An experimental group of 23 students in 5 small high schools was given instruction in Spanish for one year entirely by open circuit television. Two high school control groups, one of 31 students in 3 small high schools and one of 96 students in a large high school in a partially Spanish-speaking community, participated in the study. These control groups were given Spanish instruction for one year in the conventional manner. For additional perspective a college class of 14 students was also observed as they were given instruction in elementary Spanish according to the usual classroom routine.

The vocabulary ability of these four groups was assessed every two weeks with a 15 minute quiz. Nine parallel quizzes had been developed for this assessment from vocabulary items of the elementary Cooperative Spanish Test. Learning curves were generated analytically to fit the data

obtained from these quizzes. It was assumed that a modification of a rational hyperbola derived by Thurstone would be an appropriate model. The four parameters mentioned above were obtained by algebraic computation from each student's learning curve.

The Findings. There were no significant differences between experimental and control groups with regard to intelligence, learning rate and regularity of performance. There were significant differences with regard to level of achievement. The large school control group was more proficient at the outset; the small school control and experimental group were about the same. By the end of the first semester the small school control group had passed the large school control group but the experimental group was significantly behind both. At the end of the year, the first semester differences were magnified. The results of the standardized tests substantiated the results of analysis of the learning curves.

Conclusions. The significantly poorer achievement of the TV students is not attributable to a lack of scholastic ability, but it might have been caused by other factors than the medium of instruction. There was no attempt to equate teachers nor educational objectives in this experiment. The same instructional program had been conducted and evaluated the previous year (Rottmann, 1959) and the TV facilities for Spanish instruction and this TV teacher had been rated by the students as less desirable than conventional facilities and classroom teachers. It is certain that the teachers participating in this subsequent experiment differed as to the importance they attached to the teaching of vocabulary and as to how they would react to their students' objections to taking quizzes on material that they had not covered in class. These matters contaminate the results to a

certain extent, but the results of this experiment can be and should be added to that body of experimental literature which will be the basis for establishing the proper role of educational television.

The analytic methodology used in this study seems particularly promising for describing achievements although some of the facets of this methodology certainly need refinement. A single learning curve model is powerful enough to handle very diverse performances. Parameters of these curves can be related to meaningful characteristics of behavior. The learning rate and regularity of performance parameters were not significant contributors to the understanding of learning in this experimental situation, but no reason that they could not be was encountered.

References

Carpenter, C.R. and Greenhill, L. P. (1958) An Investigation of Closed Circuit Television for Teaching University Courses Report No 2. University Park, Pennsylvania: Penn State University

Gulliksen, H. (1934) A rational equation of the learning curve based on Thorndike's Law of Effect. J. gen. Psychol., 2, 395-434.

Herminghaus, Earl G. (1958) An Investigation of Television Teaching. St. Louis: St Louis Public Schools

Kumata, Hiedya, (1956) An Inventory of Instructional Television Research, Ann Arbor, Michigan: Educational Television and Radio Center.

Lord, F. M. (1956) The Measurement of growth. J. educ. psychol. Meas. 16, 421-437.

Macomber, F. G. (1957) Experimental Study in Instructional Procedures Report No 2. Oxford, Ohio: Miami University

Neidt, Charles O. and French, Joseph L. (1958) Reaction of High School Students to Television Teachers. Lincoln Nebraska: University of Nebraska.

Rottmann, Leon. (1959) The Effectiveness of High School Television-Correspondence Instruction Lincoln Nebraska: University of Nebraska.

Stake, Robert E. (1958) Learning Parameters, Aptitudes and Achievements, Princeton, New Jersey: Educational Testing Service, 1958.

Tanner, Sheldon C. (1957) Experiments in teaching by closed circuit television. School and Society 85 : 229-33, June 22.

Thurstone, L. L. (1930) The learning function. J. gen Psychol., 3 469-493.

Wiley, L. E. and Wiley, A. M. (1937) Studies in the learning function. Psychometrika, 2 1-9, 107-120, 161-164.

Woodrow, H. (1946) The ability to learn, Psychol. Rev., 53, 147-158.